CCSLM: Context as Parameter Assignment for Expertise

- **Central idea**
  - The main role of context is **not to determine the output**.
  - The main role of context is to **assign regions of parameter space in which expertise can grow**.
  - Context determines which parameters participate in a particular portion of the learning problem.
  - Repeated exposure to similar contexts causes those parameters to specialize.
- **From context to expertise**
  - Context selects or modulates parameters.
  - Training repeatedly adjusts the selected parameters.
  - Over time, particular parameter regions become specialized.
  - These regions can be regarded as **future experts**.
  - Each expert is fundamentally a **linear mapping and linear associative memory**.
- **Context is control, not computation**
  - Context provides a control pathway.
  - It does not itself transform the data.
  - The selected parameters perform the data transformation.
  - Consequently, the final output is entirely the result of the selected linear mappings and their concatenation.
  - There is no separate context-dependent transformation of the data.
- **Associative memory remains active within each context**
  - Selecting a parameter region does not select a predetermined answer.
  - The selected linear mapping still responds to the particular input.
  - Different inputs can therefore produce rich, different recalls within the same context.
  - The context reduces the associative problem; it does not replace the association.
- **Fractional associative memory**
  - A large associative-memory problem can be distributed across many parameter regions.
  - Each region becomes a fragment of the overall associative memory.
  - Context determines which fragment receives learning pressure and which fragment is subsequently used.
  - This can keep individual fragments below critical storage capacity while the complete system contains many more associations.
- **Expertise emerges through training**
  - The system need not be explicitly told which parameters should become experts.

    - Context repeatedly exposes different parameter regions to different portions of the training distribution.
    - Backpropagation then adjusts those regions according to the final task.
    - Expertise therefore emerges from the interaction of **contextual parameter assignment and learning pressure**.
- **Experts must cooperate**
    - The parameter regions are not independent lookup tables.
    - A mapping selected at one stage must produce representations useful to mappings selected at later stages.
    - Backpropagation coordinates the fragments through the complete concatenated computation.
    - The resulting experts can therefore become complementary parts of a larger associative computation.
- **ReLU and moving contexts**
    - In a ReLU implementation, the effective linear mapping depends on activation patterns.
    - Weight changes can therefore change the contexts themselves during learning.
    - Early in training, parameter assignments and contexts may move substantially.
    - This is a genuine complication: the system is simultaneously learning the mappings and changing the regions that select them.
    - As learning begins to stabilize, the activation patterns and corresponding contexts should also tend toward greater stability.
    - Parameter specialization can then consolidate around these increasingly stable contexts.
- **A useful conceptual sequence**
  Context → parameter assignment → repeated learning pressure → specialization → expert linear associative mapping → coordinated concatenation
- **Most concise formulation**

  Context assigns parameter regions in which expertise can grow. Training specializes those regions into expert linear associative mappings. At inference, context selects the appropriate parameters, while the selected linear mappings perform all of the data transformation and associative recall. Backpropagation coordinates the resulting memory fragments into a coherent computation.

- **Key distinction**
    - **Context does not compute the answer.**
    - **Context determines where expertise develops and which expertise is subsequently brought into play.**
    - **The linear associative mappings do the computing.**